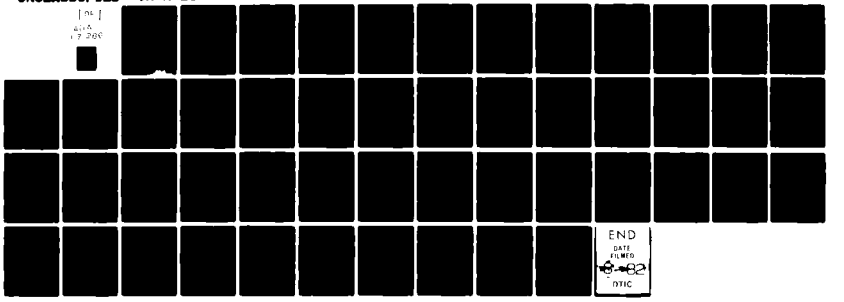
AD-A117 266 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
AUTOREGRESSIVE SPECTRAL ESTIMATION.(U)
JUN 82 E PARZEN DAAG29-80-C-0070
UNCLASSIFIED TR-N-28 NL



TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

INSTITUTE OF STATISTICS
713 - 845-3141

AUTOREGRESSIVE SPECTRAL ESTIMATION

by Emanuel Parzen

Institute of Statistics

Texas A&M University

Technical Report No. N-28

June 1982

Texas A&M Research Foundation
Project No. 4226T

"Multiple Time Series Modeling and
Time Series Theoretic Statistical Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; **distribution unlimited.**

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-28	2. GOVT ACCESSION NO. AD-A117 266	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AUTOREGRESSIVE SPECTRAL ESTIMATION		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-82-MP-2001 ARU DAAG29-80-C-0070
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Institute of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE June 1982
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Time series analysis, autoregressive spectral density, maximum entropy, empirical spectral analysis.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Autoregressive spectral estimation is one of the new techniques for spectral analysis developed in the last two decades. Its theory and applications are extremely extensive. This article aims to provide an overview (rather than a detailed account) of the main ideas. A comprehensive bibliography guides the reader to articles in which case studies and comparisons of autoregressive spectral estimators are described.

AUTOREGRESSIVE SPECTRAL ESTIMATION

by Emanuel Parzen

Institute of Statistics, Texas A&M University

0. Introduction

The problem of spectral analysis of time series is clearly of great interest to the many applied scientists who use spectral analysis in their scientific research. It should be of great interest to statisticians because it embodies prototypes of two of the great problems of modern statistics: functional inference and modeling. A problem of statistical inference usually assumes three ingredients: a sample of observations, a parameter which indexes the family of possible joint probability densities of the sample, and a formula for the probability density of the sample

$$f(\text{sample}|\text{parameter}).$$

Classical statistical inference assumes the parameter is a finite dimensional parameter $\theta = (\theta_1, \ldots, \theta_k)$. Functional inference assumes the parameter is a function, such as $f(\omega)$, $0 \leq \omega \leq 1$.

The parameter estimation problem seeks to form optimal estimators (denoted $\hat{\theta}$) of the parameter. A typical model

---

This research was supported by the Office of Naval Research (Contract N00014-81-MP-10001, ARO DAAG29-80-C0070).

identification problem seeks to find the smallest number of significantly non-zero components θ_j of the parameter θ.

Estimation of a function often has similar features to model identification, since a function can be parametrized exactly by a countable infinity of parameters. However in practice one can only efficiently estimate a finite number of parameters. Therefore to estimate a function one must use the smallest finite number of parameters which provide an adequate approximation of the function.

The goals of functional inference and model identification are in my view best pursued simultaneously. One seeks methods of statistical inference which are finite-parametric and non-(or infinite-) parametric. One achieves this goal by using finite parameter models (which theory indicates might be exact methods) in ways that enable them to be interpreted also as approximating models.

Autoregressive spectral estimation is one of the new techniques for spectral analysis developed in the last two decades. Its theory and applications are extremely extensive. This article aims to provide an overview (rather than a detailed account) of the main ideas. A comprehensive bibliography guides the reader to articles in which case studies and comparisons of autoregressive spectral estimators are described.

The spectral density is defined in section 1. Infinite order AR and MA representations of a stationary time series are introduced in section 2. Entropy as a motivation for

autoregressive schemes is discussed in section 3. Alternate parametrizations of an autoregressive scheme are outlined in section 4. Algorithms for computing the coefficients of autoregressive spectral densities are stated in section 5. Criteria for determining the orders of autoregressive schemes are mentioned in section 6. Suggestions for empirical spectral analysis are outlined in section 7. The final section provides a guide to the literature of autoregressive spectral estimation by listing the articles which correspond to some important developments.

1. Correlations and Spectral Functions of a Stationary Time Series

The theory of time series discusses separately discrete parameter time series $\{Y(t),\ t=0, \pm1, \pm2, \ldots\}$ and continuous parameter time series $\{Y(t),\ -\infty<t<\infty\}$. Only the former case is discussed in this article. Discrete parameter series usually arise by observing a continuous parameter time series at equi-spaced times. The frequency variable ω of a pure harmonic

$$Y(t) = A \cos 2\pi\omega t + B \sin 2\pi\omega t \tag{1}$$

observed at $t = 0, \pm1, \ldots$ can be assumed to vary in either $-0.5 \leq \omega \leq 0.5$ or $0 \leq \omega \leq 1$. The first interval is usually adopted, but the second interval will be adopted in this article because it is more convenient for developing

isomorphisms between spectral analysis and non-parametric data modeling using quantile and density-quantile functions [see Parzen (1979)].

The definitions and notation we adopt for the functions used to describe a zero mean stationary Gaussian discrete parameter time series $Y(t)$, $t=0, \pm 1, \ldots$ are as follows.

A "time domain" specification of the probability law of $Y(\cdot)$ is provided by the covariance function

$$R(v) = E[Y(t)Y(t+v)], \quad v=0, \pm 1, \pm 2, \ldots; \tag{2}$$

or by the variance $R(0)$ and the correlation function

$$\rho(v) = \frac{R(v)}{R(0)} = \text{Corr}\,[Y(t), Y(t+v)]. \tag{3}$$

To define spectral (frequency) domain specification of the probability law of $Y(\cdot)$ we first assume summability of $R(\cdot)$ and $\rho(\cdot)$. The Fourier transforms of $R(v)$ and $\rho(v)$ are called the power spectrum $S(\omega)$ and spectral density function $f(\omega)$ respectively, and are defined by

$$S(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v \omega} R(v), \quad 0 \leq \omega \leq 1; \tag{4}$$

$$f(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v \omega} \rho(v), \quad 0 \leq \omega \leq 1. \tag{5}$$

The spectral distribution function is defined by

$$F(\omega) = \int_0^{\omega} f(\omega')\, d\omega', \quad 0 \leq \omega \leq 1 \tag{6}$$

For data analysis one actually computes a modified spectral distribution function denoted $F_+(w)$ and defined for $0 \leq \omega \leq 0.5$:

$$(7) \quad F_+(\omega) = 2F(\omega), \quad 0 \leq \omega \leq 0.5.$$

A correlation function $\rho(v)$ has a mathematical property of being a positive definite function which, without assuming summability, guarantees the existence of: (1) a spectral distribution function $F(\omega)$, and (2) the spectral representation of the correlation function $\rho(v)$ given by

$$(8) \quad \rho(v) = \int_0^1 e^{2\pi i v\omega} \, dF(\omega)$$

$$= \int_0^{0.5} \cos 2\pi v\omega \, dF_+(\omega),$$

The notion of an ergodic time series is not given a rigorous definition in this article but its intuitive meaning is important for us. We call a time series ergodic when the parameters of its probability law possess consistent estimators (and thus can be determined with probability one, given a sample of infinite length). An example of a non-ergodic stationary Gaussian zero mean time series is

$$Y(t) = A \cos 2\pi\omega t + B \sin 2\pi\omega t$$

where A and B are independent $N(0,\sigma^2)$ random variables. One can infer the values of A and B in the sample observed, but one cannot infer the value of σ^2. This time series has correlation function

$$\rho(v) = \cos 2\pi\omega v, \quad v=0, \pm 1, \pm 2, \ldots, \tag{9}$$

and does not possess a spectral density.

Parzen (1982) proposes that it is useful in practice to distinguish qualitatively between three types of time series, which we call

no memory: white noise,

short memory: stationary and ergodic,

long memory: non-stationary or non-ergodic.

A no-memory or white noise time series is a stationary Gaussian time series satisfying either of the equivalent conditions:

$$\rho(v) = 0 \text{ for } v>0; \tag{10}$$

$$f(\omega) = 1, 0 \leq \omega \leq 1.$$

A short memory time series is a stationary time series possessing a summable correlation function $\rho(v)$ and a spectral density $f(\omega)$ which is bounded above and below in the sense that the dynamic range of $f(\omega)$

$$DR(f) = \max_{0\leq\omega\leq 1} f(\omega) \div \min_{0\leq\omega\leq 1} f(\omega) \tag{11}$$

satisfies $1< DR(f)< \infty$.

A long memory time series is one which is neither no memory nor short memory; alternatively, a long memory time

series is one which is non-stationary or non-ergodic. It usually has components representing cycles or trends. An example of a long memory time series is (1) where A and B are independent $N(0,\sigma^2)$ random variables.

For a short memory time series one can define the inverse-correlation function

$$\rho i(v) = \int_0^1 e^{2\pi i v\omega} f^{-1}(\omega)d\omega \div \int_0^1 f^{-1}(\omega)\, d\omega \tag{12}$$

and the cepstral-correlation function

$$\gamma(v) = \int_0^1 e^{2\pi i v\omega} \log f(\omega)\, d\omega \tag{13}$$

It should be noted that the inverse-correlation function is positive definite. However the cepstral-correlation function is not. These new types of correlation functions are introduced because they may provide more parsimonious parameterizations in the sense that they decay to 0 faster than does the correlation function. Statistical inference from a sample of the probability law of a time series often achieves greatest statistical efficiency by using the most parsimonious parametrizations. Thus to form estimators $\hat{f}(\omega)$ of the spectral density $f(\omega)$ from a raw estimator $\tilde{f}(\omega)$, greater precision may be attained by first forming estimators $\{f^{-1}(\omega)\}\hat{}$ and $\{\log f(\omega)\}\hat{}$ of the inverse or logarithm of the spectral density. Autoregressive spectral estimation may be regarded as an approach to estimating $f(\omega)$ by first estimating $f^{-1}(\omega)$.

2. Filter Models of a Short Memory Stationary Time Series

A short memory zero mean Gaussian stationary time series $Y(t)$ has a property of fundamental importance: it can be linearly transformed to a white noise time series, denoted $Y^{\nu}(t)$ or $\varepsilon(t)$, by a filter that is invertible. The definition of $Y^{\nu}(t)$ is provided by the theory of minimum mean square error prediction.

The definitions and notation of prediction theory that we adopt are as follows:

$$(1) \qquad Y^{\mu,m}(t) = E[Y(t)|Y(t-1),\ldots,Y(t-m)]$$

denotes the memory m one-step ahead predictor while

$$(2) \qquad Y^{\nu,m}(t) = Y(t) - Y^{\mu,m}(t)$$

is the prediction error, and

$$(3) \qquad \sigma_m^2 = E[|Y^{\nu,m}(t)|^2] + E[|Y(t)|^2]$$

is the mean square prediction error measured in units of the variance $R(0)$ of $Y(t)$. Corresponding infinite memory notation is

$$(4) \qquad Y^{\mu}(t) = E[Y(t)|Y(t-1), Y(t-2),\ldots]$$

$$(5) \qquad Y^{\nu}(t) = Y(t) - Y^{\mu}(t)$$

$$(6) \qquad \sigma_{\infty}^2 = E[|Y^{\nu}(t)|^2] + E[|Y(t)|^2].$$

Explicit formulas for the foregoing predictors, prediction errors, and mean square prediction errors can be obtained from the correlation function $\rho(v)$. The autoregressive coefficients $\alpha_m(1),\ldots,\alpha_m(m)$ of order m are defined by

$$(7) \qquad -Y^{\mu,m}(t) = \alpha_m(1)\ Y(t-1) + \ldots + \alpha_m(m)\ Y(t-m),$$

$$(8) \qquad Y^{\nu,m}(t) = Y(t) + \alpha_m(1)\ Y(t-1) + \ldots + \alpha_m(m)\ Y(t-m).$$

A predictor is characterized by the condition that the prediction error is orthogonal (normal) to the predictor variables:

$$(9) \qquad E[Y^{\nu,m}(t)Y(t-k)] = 0, \quad k=1,\ldots,m$$

By substituting (8) into (9) one obtains the famous Yule-Walker equations, defining $\alpha_m(0) = 1$.

$$(10) \qquad \sum_{j=0}^{m} \alpha_m(j)\ \rho(j-k) = 0, \quad k=1,\ldots,m$$

One obtains σ_m^2 by

$$(11) \qquad \sigma_m^2 = E[Y^{\nu,m}(t)\ Y(t)] \div E[|Y(t)|^2] = \sum_{j=0}^{m} \alpha_m(j)\ \rho(j)$$

For a short memory time series, these equations also hold with $m=\infty$.

The time series of infinite memory prediction errors $Y^{\nu}(t)$ is always a white noise series called the innovations.

It provides a transformation of the time series Y(t) to a white noise time series $Y^{\nu}(t)$ which we write

$$Y(t) + \alpha_{\infty}(1)\, Y(t-1) + \ldots = Y^{\nu}(t), \tag{12}$$

and call an AR(∞) or infinite order autoregressive scheme for Y(t). An MA(∞) or infinite order moving average scheme representation is

$$Y(t) = Y^{\nu}(t) + \beta_{\infty}(1)\, Y^{\nu}(t-1) + \ldots \tag{13}$$

whose coefficients $\beta_{\infty}(k)$ can be determined recursively from $\alpha_{\infty}(j)$ by $\beta_{\infty}(0) = 1$, and for $k>0$

$$\beta_{\infty}(k) + \sum_{j=1}^{k} \alpha_{\infty}(j)\, \beta_{\infty}(k-j) = 0 \tag{14}$$

The AR(∞) and MA(∞) representations have important implications for spectral analysis since they provide formulas for the spectral density function f(ω) alternative to the formula that f(ω) is the Fourier transform of ρ(v). One can show that

$$f(\omega) = \sigma_{\infty}^{2} |h_{\infty}(e^{2\pi i\omega})|^{2}\ , \tag{15}$$

$$f^{-1}(\omega) = \sigma_{\infty}^{-2} |g_{\infty}(e^{2\pi i\omega})|^{2}\ , \tag{16}$$

defining

$$h_{\infty}(z) = \sum_{j=0}^{\infty} \beta_{\infty}(j)\, z^{j}, \quad g_{\infty}(z) = \sum_{k=0}^{\infty} \alpha_{\infty}(k)\, z^{k}. \tag{17}$$

These infinite series converge in general in mean square on the unit interval. In order to guarantee pointwise convergence at each ω in $0 \leq \omega \leq 1$ one must make an additional smoothness condition such as a Lipshitz condition on $f(\omega)$, which is implied in turn by the condition

$$(18) \qquad \sum_{v=-\infty}^{\infty} |v||\rho(v)| < \infty \quad .$$

In spectral analysis we usually assume at least the existence of a continuous second derivative, which is implied by the condition

$$(19) \qquad \sum_{v=-\infty}^{\infty} |v|^2 |\rho(v)| < \infty.$$

The notion of a time series $Y(\cdot)$ being an autoregressive scheme of order p, denoted AR(p), can be defined in terms of prediction theory as follows: $Y(\cdot)$ is an AR(p) if and only if the memory p prediction errors $Y^{\nu,p}(\cdot)$ is a white noise time series and $\alpha_p(p) \neq 0$. The spectral density of $Y^{\nu,p}(\cdot)$ can be expressed in terms of the autoregressive transfer function of order p

$$(20) \qquad g_p(z) = \alpha_p(0) + \alpha_p(1)\, z + \ldots + \alpha_p(p)\, z^p$$

by

$$(21) \qquad f_{Y^{\nu},p}(\omega) = \frac{1}{\sigma_p^2} \, |g_p(e^{2\pi i\omega})|^2 \, f(\omega)$$

If the time series $Y(\cdot)$ is in fact AR(p), then its spectral density equals the function

$$(22) \qquad f_p(\omega) = \sigma_p^2 \, |g_p(e^{2\pi i\omega})|^{-2}$$

which we call, in general, the approximating autoregressive spectral density of order p. A time series $Y(\cdot)$ can be regarded as approximated by an AR(p) if

$$(23) \qquad \bar{f}_p(\omega) = \frac{f(\omega)}{f_p(\omega)}$$

can be regarded as not "significantly" different from the constant 1. In this way a test of the hypothesis that a time series $Y(\cdot)$ is AR(p) can be converted to a test of the hypothesis that the prediction error time series is white noise.

The sequence of approximating autoregressive spectral densities $f_m(\omega)$, $m=1,2,\ldots$ may be shown to converge as m tends to ∞ to $f(\omega)$ at each ω in $0 \leq \omega \leq 1$ under suitable conditions (see especially Nevai (1979)). Sufficient conditions are that $f(\omega)$ has finite dynamic range (and therefore is bounded above and below) and has a continuous derivative. When an estimator, denoted $\hat{f}_m(\omega)$, of $f_m(\omega)$ is used as an estimator of $f(\omega)$, one has to take into account two kinds of errors, called respectively

bias and variance. Bias is a measure of the deterministic difference between $f_m(\omega)$ and $f(\omega)$; while variance is a measure of the stochastic distance between $\hat{f}_m(\omega)$ and $f_m(\omega)$. As m increases bias decreases while variance increases. This is an example of the fundamental problem of empirical spectral naalysis which is how to achieve an optimal balance between bias and variance. When one uses autoregressive spectral estimation, this problem reduces to a question of determining the order m of the approximating autoregressive scheme, which is discussed in section 6.

It should be noted that basic references for the mathematical properties (especially convergence) of autoregressive transfer functions $g_m(z)$ are Geronimus (1960) and Grenander and Szegö (1958).

3. Entropy Interpretation of Autoregressive Spectral Densities

The use of autoregressive spectral densities as exact models, and as approximating models, for true spectral densities is often questioned by sceptical statisticians on the ground that their use in general is ad hoc and without theoretical justification. A possible answer to this criticism is provided by entropy concepts. This section reviews these concepts in order to state their application to spectral estimation.

The notion of entropy in statistics is usually first defined for a discrete distribution with probability mass function $p(x)$. The entropy of this distribution, denoted $H(p)$, is defined by

$$(1) \qquad H(p) = -\sum_{x} p(x) \log p(x)$$

For the distribution of a continuous real valued random variable X, with probability density function $f(x)$, entropy is defined (analogously or formally) by

$$(2) \qquad H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx$$

A concept closely related to entropy is information divergence $I(f;g)$ between two probability densities $f(x)$ and $g(x)$, defined by

$$(3) \qquad I(f;g) = \int_{-\infty}^{\infty} \{-\log \frac{g(x)}{f(x)}\} f(x)\, dx$$

One should note that $I(f;g)$ equals minus the generalized entropy $H(f|g)$ defined by

$$(4) \qquad H(f|g) = \int_{-\infty}^{\infty} \{-\frac{f(x)}{g(x)} \log \frac{f(x)}{g(x)}\} g(x)\, dx$$

Another fundamental concept is cross-entropy defined by

$$(5) \qquad H(f;g) = \int_{-\infty}^{\infty} \{-\log g(x)\} f(x)\, dx.$$

Note that $H(f) = H(f;f)$.

Information divergence is expressed in terms of cross-entropy and entropy by

$$(6) \qquad I(f;g) = H(f;g) - H(f)$$

Important Information Inequality:

$$(7) \qquad I(f;g) \geq 0$$

with equality if and only if $f = g$; consequently

$$(8) \qquad H(f) \leq H(f;g)$$

Proof: Apply Jensen's inequality which states for an arbitrary function $h(x)$

$$(9) \qquad \int_{-\infty}^{\infty} \{\log h(x)\} f(x)\, dx \leq \log \int_{-\infty}^{\infty} h(x) f(x)\, dx$$

with equality if and only if $h(x) = 1$ for almost all x with respect to the probability density $f(x)$.

Some applications of entropy in probability and statistical modeling are as follows.

The method of maximum likelihood parameter estimation can be described abstractly as follows. One introduces a parametric family of probability densities $f_\theta(x)$, indexed by a vector parameter $\theta = (\theta_1, \ldots, \theta_k)$. Suppose there is a true parameter value $\bar{\theta}$ in the sense that the true probability density $f(x) = f_{\bar{\theta}}(x)$. Then $\bar{\theta}$ satisfies

$$H(f) = H(f; f_{\bar{\theta}}) = \min_\theta H(f; f_\theta). \tag{10}$$

To estimate $\bar{\theta}$ from data, one forms an estimator $\tilde{H}(f; f_\theta)$ of $H(f; f_\theta)$ and defines an estimator $\hat{\theta}$ of $\bar{\theta}$ by

$$\tilde{H}(f; f_{\hat{\theta}}) = \min_\theta \tilde{H}(f; f_\theta). \tag{11}$$

The estimator $\tilde{H}(f; f_\theta)$ could be of the form

$$\tilde{H}(f; f_\theta) = H(\tilde{f}; f_\theta) \tag{12}$$

for a suitable raw estimator $\tilde{f}(x)$ of $f(x)$.

The parametric families of probability densities $f_\theta(x)$ are often derived axiomatically using a maximum entropy principle.

Theorem: Fix k functions $T_j(x)$, $j=1,2,\ldots,k$, and k real numbers $\tau_1, \tau_2, \ldots, \tau_k$ such that there exists probability densities f(x) satisfying

$$\int_{-\infty}^{\infty} T_j(x)\, f(x)\, dx = \tau_j, \quad j=1,\ldots,k. \tag{13}$$

The density with maximum entropy H(f) among these densities is of the form

$$\log f_\theta(x) = \sum_{j=1}^{k} \theta_j T_j(x) - \Psi(\theta_1,\ldots,\theta_k) \tag{14}$$

where

$$\Psi(\theta_1,\ldots,\theta_k) = \log \int_{-\infty}^{\infty} dx \exp \sum_{j=1}^{k} \theta_j T_j(x) \quad , \tag{15}$$

and $\theta_1,\ldots,\theta_k$ are chosen to satisfy

$$\int_{-\infty}^{\infty} T_j(x)\, f_\theta(x)\, dx = \tau_j \quad , \quad j=1,\ldots,k. \tag{16}$$

Proof: The theorem may be proved by calculus of variations arguments. A quick proof is to verify that for any f(x) satisfying the moment constraints (13)

$$H(f;\, f_\theta) = \Psi(\theta_1,\ldots,\theta_k) - \sum_{j=1}^{k} \theta_j \tau_j = H(f_\theta), \tag{17}$$

and therefore

$$(18) \qquad H(f) \leq H(f;f_\theta) = H(f_\theta).$$

Thus the maximum entropy is achieved by $f_\theta(x)$.

Natural Exponential models. A parametric family of probability densities $f_\theta(x)$ is said to obey a natural exponential model when it is of the form (14). Thus natural exponential models are maximum entropy probability densities.

To extend entropy concepts to short memory stationary zero mean Gaussian time series, define the information divergence for a sample $\{Y(t),\ t=1,2,\ldots,T\}$ as a function of the true probability density f of the sample, and a model g for f. We define

$$(19) \qquad I(f;\ g) = \lim_{T\to\infty} I_T(f;g) \quad ,$$

$$(20) \qquad I_T(f;g) = \frac{-1}{T} E_f \left[\log \frac{g(Y(1),\ldots,Y(T))}{f(Y(1),\ldots,Y(T))} \right]$$

It should be noted that we are using the notation f and g with a variety of meanings. For a Gaussian zero mean stationary time series, the probability density of the sample is specified by the spectral densities $f(\omega)$ of the true distribution and $g(\omega)$ of the model. The arguments of the information divergence I(f;g) indicate spectral densities in the following discussion. Pinsker (1963) derives the following very important formula:

$$(21) \qquad I(f;g) = \frac{1}{2} \int_0^1 \left\{ \frac{f(\omega)}{g(\omega)} - \log \frac{f(\omega)}{g(\omega)} - 1 \right\} d\omega$$

Since $u - \log u - 1 \geq 0$ for all u, I has two of the properties of a distance: $I(f;g) \geq 0$, $I(f;f) = 0$. However I does not satisfy the triangle-inequality.

We define the cross-entropy of spectral density functions $f(\omega)$ and $g(\omega)$ by

$$(22) \qquad H(f;g) = \frac{1}{2} \int_0^1 \{\log g(\omega) + \frac{f(\omega)}{g(\omega)}\} \, d\omega$$

The entropy of f is

$$(23) \qquad H(f) = H(f;f) = \frac{1}{2} \int_0^1 \{\log f(\omega) + 1\} \, d\omega$$

Information divergence can be expressed

$$(24) \qquad I(f;g) = H(f;g) - H(f)$$

Hence

$$(25) \qquad H(f) \leq H(f;g)$$

An approximating autoregressive spectral density of order m, denoted $\bar{f}_m(\omega)$, to a spectral density $f(\omega)$ is defined by

$$(26) \qquad H(f;\bar{f}_m) = \min_{f_m} H(f;f_m)$$

where the minimization is over all f_m of the form

$$(27) \qquad f_m(\omega) = \sigma_m^2 \, |g_m(e^{2\pi i\omega})|^{-2} \, ,$$

$$(28) \qquad g_m(z) = 1 + \alpha_m(1)\, z + \ldots + \alpha_m(m)\, z^m$$

One may verify that

$$(29) \qquad H(f;f_m) = \frac{1}{2} \{\log \sigma_m^2 + \frac{1}{\sigma_m^2} \int_0^1 g_m|(e^{2\pi i\omega})|^2 \, f(\omega)\, d\omega\}$$

The coefficients $\bar{\sigma}_m^2$, $\bar{\alpha}_m(1),\ldots,\bar{\alpha}_m(m)$ of the minimum cross-entropy approximating autoregressive spectral density satisfy

$$(30) \qquad \bar{\sigma}_m^2 = \int_0^1 |\bar{g}_m(e^{2\pi i\omega})|^2 \, f(\omega) \, d\omega$$

$$= \int_0^1 \bar{g}_m(e^{2\pi i\omega}) \, f(\omega) \, d\omega = \sum_{j=0}^{m} \hat{\alpha}_m(j) \, \rho(j)$$

$$= \min_{g_m} \int_0^1 |g_m \, (e^{2\pi i\omega})|^2 \, f(\omega) \, d\omega,$$

$$(31) \qquad \int_0^1 \bar{g}_m(e^{2\pi i\omega}) e^{-2\pi ik\omega} \, f(\omega) \, d\omega$$

$$= \sum_{j=0}^{m} \bar{\alpha}_m(j) \, \rho(j-k) = 0, \quad k=1,2,\ldots,m$$

Further

$$(32) \qquad H(f;\bar{f}_m) = \frac{1}{2} \{\log \bar{\sigma}_m^2 + 1\} = H(\bar{f}_m)$$

The autoregressive spectral density $\bar{f}_m(\omega)$ can be derived axiomatically using a maximum entropy principle.

Theorem: The spectral density with maximum entropy among all spectral densities $f(\omega)$ satisfying the constraints

$$(33) \qquad \int_0^1 e^{2\pi i\omega j} f(\omega)\, d\omega = \rho(j), \quad j=1,2,\ldots,m$$

for m specified correlation coefficients $\rho(1),\ldots,\rho(m)$ is $\bar{f}_m(\omega)$

Proof: It may be verified that $\bar{f}_m(\omega)$ satisfies the constraints (33), and (32) holds for any $f(\omega)$ satisfying (33). Since

$$(34) \qquad H(f) \leq H(f;\bar{f}_m) = H(\bar{f}_m),$$

it follows that $\bar{f}_m$ has maximum entropy among all spectral densities satisfying the constraints (33).

The maximum entropy principle provides a motivation or justification for the use of autoregressive spectral estimators. However the maximum entropy principle provides no insight into how to identify an optimal order m, or even what are the effects of different methods of estimating the parameters σ_m^2, $\alpha_m(1),\ldots,\alpha_m(m)$. It provides no guidance for how to learn from the data whether the time series is non-stationary (long memory) or stationary (short memory), or whether the best time series model is AR, MA, or ARMA. It is a principle for deriving probability models, rather than statistically fitting models to data. Further, the maximum entropy principle justifies autoregressive estimators only for short memory time series. Autoregressive estimators are justified for long memory time series by the fact that a pure harmonic $Y(t) = A \cos 2\pi\omega t + B \sin 2\pi\omega t$ satisfies $Y(t)-\phi Y(t-1)+Y(t-2)=0$ where $\phi = 2 \cos 2\pi\omega$.

4. Parametrizations of Autoregressive Spectral Estimators

There are many ways for forming autoregressive spectral estimators, because there are four equivalent ways of parametrizing them.

A. Autoregressive coefficients. Consider coefficients $\alpha_m(1),\ldots,\alpha_m(m)$ such that $g(z) = 1 + \alpha_m(1)\, z + \ldots + \alpha_m(m) z^m$ satisfies $g(z) \neq 0$ for complex z such that $|z| \leq 1$. We call $g(z)$ a minimum phase filter transfer function. The autoregressive coefficients determine σ_m^2 by

$$1 = \sigma_m^2 \int_0^1 |g_m(e^{2\pi i\omega})|^{-2} d\omega$$

One computes the correlation coefficients $\rho(1),\ldots,\rho(m)$ by

$$\rho(j) = \int_0^1 \exp(2\pi i\omega j)\, \sigma_m^2\, |g_m(e^{2\pi i\omega})|^{-2}\, d\omega.$$

B. Correlations. A set of m coefficients $\rho(1),\ldots,\rho(m)$ such that the matrix [with $\rho(0) = 1$, $\rho(-v) = \rho(v)$]

$$\begin{bmatrix} \rho(0) & \rho(-1) & \ldots & \rho(1-m) \\ \rho(1) & \rho(0) & \ldots & \rho(2-m) \\ & & & \\ \rho(m-1) & \rho(m-2) & \ldots & \rho(0) \end{bmatrix}$$

is positive definite are correlation coefficients of a time series. They determine autoregressive coefficients by solving the Yule

Walker equations [with $\alpha_m(0) = 1$]

$$\sum_{j=0}^{m} \alpha_m(j)\ \rho(j-k) = 0, \quad k=1,\ldots,m.$$

One computes σ_m^2 by $\sigma_m^2 = \sum_{j=0}^{m} \alpha_m(j)\ \rho(j)$.

C. Partial Correlations Consider coefficients $\pi(1),\ldots,\pi(m)$ satisfying $|\pi(j)| < 1$, $j=1,2,\ldots,m$. They represent partial correlation coefficients defined theoretically by:

$$\pi(j) = \text{Corr}\ [Y(t),\ Y(t-j)|Y(t-1),\ldots,Y(t-j+1)]$$

Partial correlation coefficients determine autoregressive coefficients and residual variances by a recursive algorithm called the Levinson-Durbin recursion [see Levinson (1947) and Durbin (1960)]: for k=1

$$\alpha_1(1) = -\pi(1), \quad \sigma_1^2 = 1 - \pi^2(1),$$

while for $k=2, 3,\ldots,m$

$$\alpha_k(k) = -\pi(k) = \frac{-1}{\sigma_{k-1}^2} \sum_{j=0}^{k-1} \alpha_{k-1}(j)\ \rho(k-j)$$

$$\sigma_k^2 = \sigma_{k-1}^2\ \{1-\pi^2(k)\}$$

$$\alpha_k(j) = \alpha_{k-1}(j) - \pi(k)\ \alpha_{k-1}(k-j)$$

Autoregressive coefficients determine partial correlation coefficients by the recursion [see Barndorff-Nielsen and Schon (1973)].

$$\alpha_{k-1}(j) = \{1-\pi^2(k)\}^{-1} \{\alpha_k(j) + \pi(k)\ \alpha_k(k-j)\}$$

D. Residual variances. Consider coefficients $\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2$ satisfying

$$1 > \sigma_1^2 > \sigma_2^2 > \ldots > \sigma_m^2 > 0 \quad ,$$

and m coefficients representing sign $\pi(1), \ldots,$ sign $\pi(m)$. The σ's represent residual variances; they determine partial correlation coefficients by a formula noted by Dickenson (1978)

$$\pi(k) = \text{sign } \pi(k) \left[1 - \frac{\sigma_k^2}{\sigma_{k-1}^2} \right]^{\frac{1}{2}}$$

5. Empirical Autoregressive Spectral Estimation

Given a sample $\{Y(t),\ t=1,2,\ldots,T\}$ of a zero mean Gaussian stationary time series whose spectral density $f(\omega)$ is to be estimated by an autoregressive spectral density estimator

$$\hat{f}_m(\omega) = \hat{\sigma}_m^2 \ |\hat{g}_m(e^{2\pi i\omega})|^{-2} \quad ,$$

$$\hat{g}_m(z) = 1 + \hat{\alpha}_m(1)\ z + \ldots + \alpha_m(m)\ z^m \quad ,$$

we define the order identification problem to be the choice of order m, and the parameter estimation problem to be the choice of algorithm for computing the coefficients $\hat{\alpha}_m(1),\ldots,\hat{\alpha}_m(m)$ and the residual variance $\hat{\sigma}_m^2$.

For a sample $Y(1),\ldots,Y(T)$ of a zero mean Gaussian stationary time series, an approximation for the joint probability density function $f_\theta(Y(1),\ldots,Y(T))$ indexed by a parameter θ is obtained as follows. We assume that the time series $Y(t)$ has been divided by $\{R(0)\}^{\frac{1}{2}}$ so that its covariance function equals its correlation function. Then

$$-2 \log f_\theta(Y(1),\ldots,Y(T)) = \log (2\pi)^T \det K_\theta + Y_T^* K^{-1} Y_T$$

where * denotes complex conjugate transpose, $Y_T^* = (Y(1),\ldots,Y(T))$, $K_\theta = E\ Y_T\ Y_T^*$ is a covariance matrix with (s,t)-element equal to $\rho_\theta(s-t)$. The subscript θ on $\rho_\theta(v)$ and $f_\theta(\omega)$ indicate that they are functions of the parameters θ, which are to be estimated.

The covariance matrix K_θ is a Toepliz matrix. Asymptotically, as T tends to ∞, all T by T Toeplitz matrices have the same eigenvectors exp $(-2\pi itj/T)$, j=0,1,..., T-1. The corresponding eigenvalues are $f_\theta(j/T)$. An approximation for likelihood function frequently adopted is therefore

$$-\frac{1}{T} \log f_\theta(Y(1),\ldots,Y(T))$$

$$= \frac{1}{2} \log 2\pi + \frac{1}{2} \int_0^1 \{\log f_\theta(\omega) + \frac{\tilde{f}(\omega)}{f_\theta(\omega)}\} \, d\omega$$

$$= \frac{1}{2} \log 2\pi + H(\tilde{f}; f_\theta)$$

where $\tilde{f}(\omega)$ is the sample spectral density defined by

$$\tilde{f}(\omega) = \sum_{t=1}^{T} |Y(t) \exp(-2\pi it\omega)|^2 \div \sum_{t=1}^{T} Y^2(t)$$

Maximum likelihood estimators $\hat{\theta}$ are asymptotically equivalent to the estimators $\hat{\theta}$ minimizing the sample cross-entropy $H(\tilde{f}; f_\theta)$.

If the parametric model $f_\theta(\omega)$ is assumed to be the spectral density of an AR(p), then estimators $\hat{\sigma}_p^2$, $\hat{\alpha}_p(1),\ldots,\hat{\alpha}_p(p)$ of the coefficients satisfy Yule-Walker equations corresponding to the sample correlation function

$$\hat{\rho}(v) = \int_0^1 e^{2\pi i \omega v} \tilde{f}(\omega)\, d\omega$$

$$= \sum_{t=1}^{T-v} Y(t)\, Y(t+v) \div \sum_{t=1}^{T} Y^2(t).$$

The sample correlation function $\hat{\rho}(v)$ can be computed, using the Fast Fourier transform, by

$$\hat{\rho}(v) = \frac{1}{Q} \sum_{k=0}^{Q-1} \exp\left(2\pi i \frac{k}{Q} v\right) \tilde{f}\left(\frac{k}{Q}\right)$$

which holds for $0 \leq v < Q - T$.

Ic should be noted that we are assuming the time series $Y(t)$ to be zero mean, or more generally to have been detrended by subtraction of $\hat{\mu}(t)$, an estimator of the mean value function $\mu(t) = E[Y(t)]$. When $\mu(t) = \mu$, a constant, we take $\hat{\mu} = \bar{Y}$. When $\mu(t)$ is a function with period d (as might be the case with d=12 for monthly time series) one might take for $\hat{\mu}(t)$ the mean of $Y(s)$ for all s=t modulo d.

By recursively solving the Yule-Walker equations one can determine sequences of (1) estimated residual variances

$$1 > \hat{\sigma}_1^2 > \hat{\sigma}_2^2 > \ldots > \hat{\sigma}_m^2 \quad \ldots,$$

(2) estimated partial correlations

$$\hat{\pi}(1),\ \hat{\pi}(2), \ldots, \hat{\pi}(m), \ldots$$

(3) estimated autoregressive coefficients

$$\hat{\alpha}_m(0) = 1, \hat{\alpha}_m(1), \ldots, \hat{\alpha}_m(m).$$

(4) autoregressive spectral density estimators

$$\hat{f}_m(\omega) = \hat{\sigma}_m^2 \left| \sum_{j=0}^{m} \hat{\alpha}_m(j) \exp 2\pi i j \omega \right|^{-2},$$

(5) residual spectral densities

$$\tilde{f}_m(\omega) = \frac{\tilde{f}(\omega)}{\hat{f}_m(\omega)} .$$

By forming a smoothed version $\hat{\tilde{f}}_m(\omega)$ of $\tilde{f}_m(\omega)$ one can obtain a final estimator $\hat{f}(\omega)$ of the unknown spectral density:

$$\hat{f}(\omega) = \hat{f}_m(\omega) \hat{\tilde{f}}_m(\omega)$$

When $\tilde{f}(\omega)$ is tested for white noise, and found not to be significantly different from white noise, then

$$\hat{f}(\omega) = \hat{f}_m(\omega),$$

and the autoregressive spectral density estimator is the final estimator.

The important question of criteria for choosing the orders of approximating spectral densities is discussed in the next section.

Computing estimators of autoregressive coefficients by solving Yule-Walker equations is called <u>stationary autoregression</u> because the autoregressive coefficients obtained are guaranteed to correspond to a stationary time series. When $\hat{\sigma}_m^2$ in the foregoing analysis is tending to approximate 0, we consider the time series to be long memory; experimental evidence indicates that more reliable estimators of the spectral density, and also of the autoregressive coefficients, are provided by <u>least-squares autoregression</u>, which solves the normal equations

$$\begin{bmatrix} \hat{K}(0,0) & \dots & \hat{K}(0,m) \\ \hat{K}(1,0) & \dots & \hat{K}(1,m) \\ \dots & \dots & \dots \\ \hat{K}(m,0) & \dots & \hat{K}(m,m) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{\alpha}_m(1) \\ \dots \\ \hat{\alpha}_m(m) \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_m^2 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

for a suitable estimator $\hat{K}(i,j)$ of

$$K(i,j) = E[Y(t-i)\ Y(t-j)]$$

Possible estimators (for i, j=0,1,...,m) are: least squares forward algorithm

$$\hat{K}(i,j) = \frac{1}{T-M} \sum_{t=0}^{T-m-1} Y(t+m-i)\ Y(t+m-j)$$

or least squares forward and backward algorithm

$$\hat{K}(i,j) = \frac{1}{2(T-M)} \sum_{t=0}^{T-m-1} \{Y(t+m-i)\ Y(t+m-j) + Y(t+i)\ Y(t+j)\}$$

When several harmonics are present in the data, whose frequencies are close together, least squares autoregressive coefficient estimators are more effective than Yule-Walker autoregressive coefficient estimators in providing autoregressive spectral estimators which exhibit the split peaks one would like to see in the estimated spectral density.

An important and popular algorithm for estimation of AR coefficients was introduced by Burg in 1967 [see Burg (1967), (1968)]. For references to descriptions of Burg's algorithm, see Kay and Marple (1981).

6. Autoregressive Order Determination

The problem of determining the orders of approximating autoregressive schemes is an important example of the problem of estimating a function by using the smallest finite number of parameters which provide an adequate approximation of the function. The true spectral density is denoted $f(\omega)$ or $f_\infty(\omega)$. An approximation $\bar{f}(\omega)$ is defined by assuming a family of densities $f_{\theta_1,\ldots,\theta_m}(\omega)$ which are functions of m scalar parameters $\theta_1,\ldots,\theta_m$. The parameter values $\bar{\theta}_1,\ldots,\bar{\theta}_m$ which minimize the cross-entropy $H(f;f_{\theta_1,\ldots,\theta_m})$ define a best approximating spectral density $\bar{f}_m(\omega) = f_{\bar{\theta}_1,\ldots,\bar{\theta}_m}(\omega)$. An estimator of f_m is $\hat{f}_m(\omega) = f_{\hat{\theta}_1,\ldots,\hat{\theta}_m}(\omega)$ where $\hat{\theta}_1,\ldots,\hat{\theta}_m$ minimizes $H(\tilde{f};f_{\theta_1,\ldots,\theta_m})$.

To evaluate the properties of $\hat{f}_m(\omega)$ as an estimator of $f_\infty(\omega)$, one must distinguish two kinds of error. The model approximation or bias error is

$$B(m) = I(f_\infty;\ \bar{f}_m).$$

The parameter estimation error or variance is

$$V(m,\ T) = E\ I\ (\bar{f}_m;\hat{f}_m)$$

As m tends to ∞, B(m) tends to 0 and V(m, T) tends to ∞. The optimal value $\hat{m}$ minimizes $EI(f_\infty;\ \hat{f}_m)$ as measured by

$$C(m) = B(m) + V(m,T)$$

In practice, one forms an estimator $\hat{C}(m)$ of the terms in $C(m)$ which depend on m.

One calls $\hat{C}(m)$ a criterion function for order determination. It should be plotted, and interpreted as a function, not just examined for its minimum value. It is useful to define a best value of m (at which $\hat{C}(m)$ is minimized) and a second best value of m (at which $\hat{C}(m)$ achieves its lowest relative minimum).

One also has to define a value $\hat{C}(0)$ of the criterion function at m=0. If

$$\hat{C}(m) > \hat{C}(0) \qquad \text{for } m=1,2,\ldots$$

then the optimum order is 0, and the time series is considered to be not significantly different from white noise. Further research is required on the properties of order determining criteria as tests for white noise.

Tests for white noise provide an alternative approach to order determination since an autoregressive estimator $\hat{f}_m(\omega)$ is regarded as an adequate fit (or smoother) if the residual spectral density $\tilde{f}(\omega) \div \hat{f}_m(\omega)$ is not significantly different from the sample spectral density of white noise.

A widely used order determining criterion is that introduced by Akaike [see Akaike (1974)]. It should be

emphasized that Akaike's criterion had a different conceptual basis than the one outlined above; it seeks to determine the order of an exact autoregressive scheme which the time series is assumed to obey. Then one can raise the objection against it that it does not consistently estimate the order, which is done by a criterion due to Hannan and Quinn (1979). Our point of view is that the approximating autoregressive scheme need only have the property that $\tilde{f}(\omega) \div \hat{f}_m(\omega)$ is just barely not significantly different from the sample spectral density of white noise.

Akaike's order determining criterion AIC is defined by

$$AIC\ (m) = \log \hat{\sigma}_m^2 + \frac{2m}{T} \quad , \ m \geq 1 \quad .$$

Possible definitions for AIC(0) are 0 or -1/T.
The Hannan and Quinn criterion is

$$AICHQ(m) = \log \hat{\sigma}_m^2 + \frac{m}{T} \log \log T \ .$$

Parzen (1974), (1977) introduced an approximating autoregressive order criterion called CAT (criterion autoregressive transfer function), defined by

$$CAT(m) = \frac{1}{T} \sum_{j=1}^{m} (1-\tfrac{j}{T})\ \hat{\sigma}_j^{-2} - (1-\tfrac{m}{T})\ \hat{\sigma}_m^{-2} \ , \ m \geq 1$$

$$CAT(0) = \ - (1 + \tfrac{1}{T}).$$

In practice CAT and AIC lead in many examples to exactly the same orders. It appears reassuring that quite different conceptual foundations can lead to similar conclusions in practice.

7. Suggestions for Empirical Spectral Analysis

The basic aim of spectral analysis is to obtain an estimated spectral density which does not introduce spurious spectral peaks, and resolves close spectral peaks. To arrive at the final form of spectral estimator in an applied problem, autoregressive spectral estimators can be used to identify the memory type of a time series (long, short, or no memory) and the type of the whitening filter of a short memory time series (AR, MA, or ARMA). An empirical time series spectral analysis should involve the following stages.

A. Pre-processing. To analyze a time series sample $Y(t)$, $t=1,\ldots,T$, one will proceed in stages which often involve the subtraction of or elimination of strong effects in order to see more clearly weaker patterns in the time series structure. The aim of pre-processing is to transform $Y(\cdot)$ to a new time series $\tilde{Y}(\cdot)$ which is short memory. Some basic pre-processing operations are memory less transformation (such as square root and logarithm), detrending, "high pass" filtering, and differencing. One usually subtracts out the sample mean $\bar{Y} = \frac{1}{T} \sum_{t=1}^{T} Y(t)$; then the time series actually processed is $Y(t) - \bar{Y}$.

B. Sample Fourier Transform by Data Windowing, Extending with Zeroes, and Fast Fourier Transform. Let $Y(t)$ denote a pre-processed time series. The first step in the analysis could be to compute successive autoregressive schemes using operations only in the time domain. An alternative first step is the computation of the sample Fourier transform

$$(1) \qquad \tilde{\psi}(\omega) = \sum_{t=1}^{T} Y(t) \exp(-2\pi i \omega t) \quad , \ \omega = \frac{k}{Q} \ , \ k = 0, 1, \ldots, Q-1,$$

at an equi-spaced grid of frequencies in $0 \leq \omega \leq 1$. We call Q the spectral computation number. One should always choose $Q \geq T$, and we recommend $Q \geq 2T$. Prior to computing $\tilde{\psi}(\omega)$, one should extend the length of the time series by adding zeroes to it. Then $\tilde{\psi}(\omega)$ can be computed using the Fast Fourier transform.

If the time series may be long memory one should compute in addition a sample "tapered" or "data windowed" Fourier transform

$$(2) \qquad \tilde{\psi}_W(\omega) = \sum_{t=1}^{T} Y(t) W(\tfrac{t}{T}) \exp(-2\pi i \omega t) \quad .$$

C. Sample Spectral Density. The sample spectral density $\tilde{f}(\omega)$ is obtained essentially by squaring and normalizing the sample Fourier transform;

$$(3) \qquad \tilde{f}(\omega) = \frac{|\tilde{\psi}(\omega)|^2}{\frac{1}{Q} \sum_{k=0}^{Q-1} |\tilde{\psi}(\tfrac{k}{Q})|^2} \quad , \ \omega = \frac{k}{Q}, \ k = 0, 1, \ldots, Q-1 \quad .$$

D. Sample Correlation Function. The sample correlation function $\hat{\rho}(v)$ is computed (using the Fast Fourier Transform).

E. Autoregressive analysis. The Yule-Walker equations are solved to estimate innovation variances $\hat{\sigma}_m^2$, to which are applied order determining criteria (AIC, CAT) to determine optimal orders $\hat{m}$ and also to test for white noise. The value of $\hat{\sigma}_{\hat{m}}^2$ and

the dynamic range of the autoregressive spectral estimator $\hat{f}_{\hat{m}}(\omega)$ are used to determine the memory type of the time series. Two orders (called the best $\hat{m}$ and second best $\hat{m}$ (2)) are determined as candidates as optimal orders corresponding to the absolute minimum and lowest relative minimum of the criterion function.

F. ARMA analysis. When a time series is classified as short memory, an approximating AR scheme of order $4\hat{m}$ is inverted to form MA (∞) coefficients which are used to estimate covariance matrix of $Y(t-j)$ and $Y^{\nu}(t-k)$. A subset regression procedure is then used to determine a "best fitting" ARMA scheme, and the corresponding ARMA spectral density estimator. One will be able to identify moving average schemes and ARMA schemes which are barely invertible, and require a long AR scheme for adequate approximation. The long AR spectral estimator introduces spurious spectral peaks when compared to the MA or ARMA estimator.

G. Non-stationary autoregression. When a time series is classified as long memory, more accurate estimators of autoregressive coefficients are provided by minimizing a least squares criterion or by Burg estimators. When several harmonics are present in the data, whose frequencies are close together, least squares autoregressive coefficient estimators are more effective than Yule-Walker autoregressive coefficient estimators in providing autoregressive spectral estimators which exhibit the split peaks one would like to see in the estimated spectral density.

H. Long Memory analysis. In the long memory case, one may want to represent Y(t) as S(t) + N(t), a long memory signal plus short memory noise. An approach to this problem may be provided by treating the sample spectral density values $\tilde{f}(k/Q)$ as a data batch to be studied by non-parametric data modeling methods using quantile functions [see Parzen (1979)]. The details of such methods are under development.

I. Nonparametric kernel spectral density estimator. An estimator $\hat{f}(\omega)$ of the spectral density is called: parametric when it corresponds to a parametric model for the time series (such as an AR or ARMA model); non-parametric otherwise. A general form of non-parametric estimator is the kernel estimator

$$\hat{f}(\omega) = \sum_{v=-\infty}^{\infty} k(\tfrac{v}{M})\ \hat{\rho}(v)\ \ e^{-2\pi i\omega v}\ \ , \quad 0 \leq \omega \leq 1.$$

The problem of determining optimum truncations points M has no general solution; one approach is to choose $M = 4\hat{m}$ to obtain a preliminary smoothing of the sample spectral density.

J. Inverse correlations and cepstral correlations. Estimators of $\rho i(v)$ and $\gamma(v)$ are computed and used to form nonparametric kernel estimators of $f^{-1}(\omega)$ and log $f(\omega)$, which may provide additional insights into the peaks and troughs to be given significance in the final estimator of the spectrum.

Extensive comparisons of different methods of spectral estimation are given in Pagano (1980), Priestley and Beamish (1981), Kay and Marple (1981). It seems clear that autoregressive

spectral estimators can give superior results when properly used. One should: determine two best orders; compute autoregressive coefficients by Yule-Walker equations and by least squares since when the time series is long memory autoregressive spectral estimators are most accurate when based on least squares estimators of autoregressive coefficients; use approximating autoregressive schemes to determine if an ARMA scheme fits better.

The end of the story of the search for the perfect spectral estimator seems attainable if one does not think of spectral estimation as a non-parametric procedure which can be conducted independently of model identification.

8. A Bibliography of Autoregressive Spectral Estimation

The bibliography aims to provide a comprehensive list of the publications in English which are directly concerned with developing the theory and methods of autoregressive spectral esatimation.

This section lists some of the publications which contributed to the development of AR spectral estimation.

Yule (1927) introduces autoregressive schemes to model disturbed periodicities as an alternative to Schuster periodogram analysis and its spurious periodicities; Yule-Walker (1931) equations relate autoregressive coefficients and correlations of a stationary time series.

Wold (1938) introduces infinite order autoregressive and moving average representations of a stationary time series; rigorous conditions are given by Akutowicz (1957).

Mann and Wald (1943) derive asymptotic distribution of estimators of autoregressive coefficients.

Levinson (1947) — Durbin (1960) derive recursive methods of solving Yule-Walker equations which subsequently lead to fast algorithms for calculation of high order AR schemes.

Whittle (1954) seems to be the first to use autoregressive schemes to estimate a spectral density. He used a low order model in a case where high order models are indicated by order determining criterion [Akaike (1974), p. 720].

Grenander and Rosenblatt (1956) criticize attempts to apply low order autoregressive schemes, and develop theory of non-parametric spectral density estimation, as do Bartlett, Parzen, and Tukey and Blackman.

Parzen (1964), Schaerf (1964), and Parzen (1968) discuss autoregressive spectral estimation as a method for empirical time series analysis; no theory is given.

Burg (1967), (1968) publishes his pioneering work on MEM (maximum entropy method of spectral estimation) and his method of calculating their coefficients.

Akaike (1969), (1970) derives asymptotic variance formulas for autoregressive spectral estimators, and states FPE (final prediction error) criterion for order determination; precursor of FPE in Davisson (1965).

Parzen (1969) derives heuristically a formula for the asymptotic variance of AR spectral estimators, confirmed by Kromer (1969) and Berk (1974); an order determining criterion is proposed.

Kromer (1969) in an unpublished Ph.D. thesis presents first rigorous analysis of asymptotic distribution of autoregressive spectral estimators, especially their bias; consistency is proved only in an iterated limit mode of convergence.

Berk (1974) provides first proof of consistency of autoregressive spectral estimators.

Carmichael (1976) in an unpublished Ph.D. thesis provides alternative proof of consistency of autoregressive estimators, and extends technique to general problems of density estimation.

Akaike (1973), (1974), (1977) introduces AIC for model order criterion and relates it to entropy maximization principles.

Parzen (1974), (1977) introduces CAT for AR order determination based on concept of finite parameter AR schemes as approximations to infinite parameter AR schemes.

Hannan and Quinn (1979) derive a modification of AIC which provides consistent estimators of the AR order, when exact model is assumed to be a finite order AR.

Huzii (1977), Shibata (1977), and Bhansali (1980), discuss rigorously the convergence of AR spectral estimators and inverse correlations.

Childers (1978) and Haykin (1979) contain very useful collections of papers.

Pagano (1980), Beamish and Priestley (1981), and Kay and Marple (1981) provide illuminating reviews of AR spectral estimators and comparisons with alternative methods.

BIBLIOGRAPHY

Akaike, H. (1969) Fitting autoregressive models for prediction, Ann. Inst. Statist. Math., 21, 243-247.

_____. (1969). Power spectrum estimation through autoregression model fitting, Ann. Inst. Statist. Math., 21, 407-419.

_____. (1970) A fundamental relation between predictor identification and power spectrum estimation. Ann. Inst. Statist. Math. 22, 219-223.

_____. (1970) Statistical predictor identification, Ann. Inst. Statist. Math., 22, 203-217.

_____. (1973) Information theory and an extension of the maximum likelihood principle. 2nd Int. Symp. on Information Theory. B. N. Petrov and F. Saki, ed. Budapest: Akademia Kiado, 267-81.

_____. (1974) A new look at the statistical model identification, IEEE Trans. Autom. Contr., AC-19, 716-723.

_____. (1977) On entropy maximization principle, Applications of Statistics, P. R. Krishnaiah, ed., North Holland, Amsterdam, 27-41.

_____. (1979) "A Bayesian Extension of the Minimum AIC procedure of Autoregressive Model Fitting, Biometrika, 66, 237-242.

Akutowicz, E. J. (1957) On an explicit formula in linear least squares prediction. Math. Scand. 5, 261-266.

Baggeroer, A. B. (1976) Confidence Intervals for Regression (MEM) Spectral Estimates, IEEE Trans. Inform. Theory, IT-22, 534-545.

Barndorf-Nielsen, O. and Schon, G. (1973) On the parametrization of autoregressive models by partial autocorrelations, J. Multivariate Analysis, 3, 408-419.

Beamish, M. and Priestley, M. B. (1981) A study of autoregressive and window spectral estimation. Appl. Statist., 30, 41-58.

Berk, K. N. (1974) Consistent autoregressive spectral estimates. Ann. Statist., 2, 489-502.

Bhansali, R. J. (1980) Autoregressive and window estimates of the inverse correlation function, Biometrika, 67, 551-66.

Bhansali, R. J. and Downham, D. Y. (1977) Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion. Biometrika 64, 547-51.

Brillinger, D. R. (1981). Time Series: Data Analysis and Theory. Expanded Edition, Holden Day: San Francisco [see pg. 512].

Burg, J. P. (1967) "Maximum entropy spectral analysis," in Proc. 37th Meeting Society of Exploration Geophysicists (Oklahoma City, OK), Oct. 31, 1967.

_____. (1968) A new analysis technique for time series data, NATO Advanced Study Institute on Signal Processing. Enschede, The Netherlands.

Carmichael, J. P. (1976) The Autoregressive Method, unpublished Ph.D. thesis, Statistical Science Division, State University of New York at Buffalo.

_____. (1978) Consistency of the Autoregressive Method of Density Estimation, Technical Report, Statistical Science Division, State University of New York at Buffalo.

Chatfield, C. (1979) Inverse autocorrelations, J. R. Statist. Soc. A 142, 363-77.

Childers, D. G. (1978) Modern Spectrum Analysis, New York: IEEE Press.

Cleveland, W. S. (1972) The inverse autocorrelations of a time series and their applications. Technometrics 14, 277-93.

Davisson, I. D. (1965) The prediction error of a stationary Gaussian time series of unknown covariance. IEEE Trans. on Information Theory, IT-11, 527-532.

Dickenson, Bradley W. (1978) Autoregressive estimation using residual energy ratios, IEEE Transactions on Information Theory, IT-24, 503-505.

Durbin, J. (1960) The fitting of time series models, Rev. Inst. de Stat., 28, 233-244.

Durrani, T. S. and Arslanian, A. S. (1982) Windows associated with high resolution spectral estimators. Submitted for publication.

Fuller, W. A., Hasen, D. P., and Goebel, J. J. (1981) Estimation of the parameters of stochastic difference equations. Annals of Statistics, 9, 531-543.

Geronimus, Y. L. (1960). Polynomials Orthogonal on a Circle and Interval, New York: Pergamon Press.

Gersch, W. (1970) Spectral analysis of EEG's by autoregressive decomposition of time series, Math. Biosci., 7, 205-222.

Gersch, W. and Sharpe, D. R. (1973) Estimation of power spectra with finite-order autoregressive models, IEEE Trans. Automat. Contr., AC-18, 367-369.

Grandell, J., Hamrud, M., and Toll, P. (1980) A remark on the correspondence between the maximum entropy method and the autoregressive models, IEEE Trans. Inform. Theory, IT-26, 750-751.

Grenander, U. and Rosenblatt, M. (1957) Statistical Analysis of Stationary Time Series, New York: Wiley.

Grenander, U. and Szego, G. (1958) Toeplitz Forms and their Applications, Berkeley: Univ. California Press.

Griffiths, L. J. and Prieto-Diaz, R. (1977) Spectral analysis of natural seismic events using auto-regressive techniques, IEEE Trans. Geosci. Electron., GE-15, 13-25.

Hannan, E. J. (1970) Multiple Time Series, Wiley: New York. [see p. 334].

Hannan, E. J. and Quinn, B. G. (1979) The determination of the order of an autoregression, Journal of the Royal Statistical Society, 41, 190-195.

Haykin, S. S., ed. (1979) Nonlinear Methods of Spectral Analysis. New York: Springer-Verlag.

Hsu, F. M. and Giordano, A. A. (1977) Line tracking using autoregressive spectral estimates, IEEE Trans. Acoustics, Speech, Signal Process., ASSP-25, 510-519.

Huzii, M. (1977) On a spectral estimate obtained by an autoregressive model fitting, Ann. Inst. Statist. Math., 29, 415-431.

Jones, R. H. (1974) Identification and autoregression spectrum estimation, IEEE Trans. Automat. Contr., AC-19, 894-898.

_____. (1975) Fitting autoregression. J. Amer. Statist. Assoc., 70, 590-592.

_____. (1976) Autoregression order selection, Geophys., 41, 771-773.

Kaveh, M. and Cooper, G. R. (1976) An empirical investigation of the properties of the autoregressive spectral estimator, IEEE Trans. Inform. Theory, IT-22, 313-323.

Kay, S. M. (1978) Improvement of autoregressive spectral estimates in the presence of noise, Rec. 1978 Int. Conf. Acoustics, Speech, and Signal Processing, 357-360.

Kay, S. M. and Marple, S. L., Jr. (1979) Sources of and remedies for spectral line splitting in autoregressive spectrum analysis, Rec. 1979 Int. Conf. Acoustics, Speech, and Signal Processing, 151-154.

Kay, S. M. (1979) The effects of noise on the auto-regressive spectral estimator, IEEE Trans. Acoust., Speech, Signal Process., ASSP-27, 478-485.

_____. (1980) Noise compensation for autoregressive spectral estimates, IEEE Trans. Acoust., Speech, Signal Process., ASSP-28, 292-303.

Kay, S. M. and Marple, Stanley L., Jr. (1981) Spectrum Analysis -- A Modern Perspective. Proceedings of the IEEE, 69, 1380-1419.

Kromer, R. E. (1969) Asymptotic properties of the auto-regressive spectral estimator. Ph.D. Thesis. Statistics Dept., Stanford University.

Lacoss, R. T. (1971) Data adaptive spectral analysis methods, Geophysics, 36. 661-675.

Landers, T. E. and Lacoss, R. T. (1977) Some geophysical applications of autoregressive spectral estimates, IEEE Trans. Geosci. Electron., GE-15, 26-32.

Levinson, N. (1946) The Wiener (root mean square) error criterion in filter design and prediction, J. Math Phys., 25, 261-278.

Mann, H. B. and Wald, A. (1943). On the statistical treatment of stochastic difference equations. Econometrica, 11, 173-220.

Marple. S. L., Jr. (1980). A new autoregressive spectrum analysis algorithm, IEEE Trans. Acoust., Speech, Signal Process., ASSP-28, 441-454.

McClave, J. (1975). Subset autoregression. Technometrics, 17, 213-220.

Morf, M., Vieira, A., Lee, D. T. L., and Kailath, T. (1978). Recursive multichannel maximum entropy spectral estimation, IEEE Trans. on Geoscience Electronics, GE-16, 85-94.

Nevai, Paul G. (1979). An asymptotic formula for the derivatives of orthogonal polynomials. SIAM J. Math. Anal. 10, 472-477.

Pagano, M. (1980). Some recent advances in autoregressive processes. Directions in Time Series, ed. D. R. Brillinger and G. C. Tiao, Institute of Mathematical Statistics, 280-302 (Comments by H. T. Davis).

Parzen, E. (1964). An approach to empirical time series, J. Res. Nat. Bur. Standards, 68B, 937-951.

_____. (1967). Time Series Analysis Papers, Holden Day: San Francisco. Includes Parzen (1964).

_____. (1967). The Role of spectral analysis in time series analysis: Review of the International Statistical Institute. 35, 125-141.

_____. (1968). Statistical spectral analysis (single channel case) in 1968, Proceedings of NATO Advance Study Institute on Signal Processing, Enschede, The Netherlands.

_____. (1969). Multiple time series modelling, Multivariate Analysis II, P. R. Krishnaiah, ed. New York: Academic Press, pp. 389-409.

_____. (1974). Some recent advances in time series modeling, IEEE Trans. Automat. Contr., AC-19, 723-730.

_____. (1977). Multiple time series: determining the order of approximating autoregressive schemes, Multivariate Analysis IV, P. Krishnaiah, ed. North-Holland: Amsterdam, 283-295.

_____. (1979). Forecasting and whitening filter estimation, TIMS Studies in the Management Sciences, 12, 149-165.

_____. (1979). Nonparametric statistical data modeling, Journal of the American Statistical Assoc., 74, 105-131.

_____. (1980). Time series modeling, spectral analysis, and forecasting, Directions in Time Series Analysis ed. D. R. Brillinger and G. C. Tiao, Institute of Mathematical Statistics.

_____. (1982). Time Series Model Identification and Prediction Variance Horizon, Applied Time Series Analysis II, ed. D. Findley, Academic Press; New York, 415-447.

Pinsker, M. (1963). Information and Information Stability of Random Variables, Holden Day: San Francisco.

Priestley, M. B. (1981). Spectral Analysis and Time Series, London: Academic Press.

Rice, J. (1979). On the estimation of the parameters of a power spectrum. J. Multivariate Analysis, 9, 378-392.

Sakai, H. (1979). Statistical properties of AR spectral analysis, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-27, 402-409.

Schaerf, Mirella C. (1964) Estimation of the covariance and autoregressive structure of a stationary time series. Ph.D. Thesis. Statistics Dept. Stanford University.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's Information Criterion. Biometrika, 63, 117-26.

_____. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Statist. 8. 147-64.

_____. (1981). An optimal autoregression spectral estimate. Annals of Statistics, 9, 300-306.

Shore, J. E. (1981). Minimum Cross-Entropy Spectral Analysis, IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-29, 230-237.

Shore, J. E. and Johnson, R. W. (1980). Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy IEEE Transactions on Information Theory, IT-26, 26-37.

Smylie, D. E., Clarke, G. K. C., and Ulrych, T. J. (1973). Analysis of irregularities in the earth's rotation, in Methods in Computational Physics, vol. 13. B. A. Bolt, ed. New York: Academic Press, pp. 391-430.

Thomson, D. J. (1977). "Spectrum estimation techniques for characterization and devleopment of WT4 waveguide-1, Bell Syst. Tech. J., 56, 1769-1815.

Tong, H. (1977). More on autoregressive model fitting with noisy data by Akaike's information criterion, IEEE Trans. Inform. Theory, IT-23, 404-410.

Ulrych, T. J. and Bishop, T. N. (1975). Maximum entropy spectral analysis and autoregressive decomposition, Rev. Geophysics Space Phys., 13, 183-200.

Ulrych, T. J. and Clayton, R. W. (1976) Time Series modelling and maximum entropy, Phys. Earth Planetary Interiors, 12, 188-200.

VanDenBos, A. (1971). Alternative interpretation of maximum entropy spectral analysis, IEEE Trans. Inform. Theory, IT-17, 493-494.

Walker, G. (1931). "On periodicity in series of related terms," Proc. Roy. Soc. London, Series A, 131, 518-532.

Whittle, P. (1954). The statistical analysis of a seiche record. J. Marine Res. 13, 76-100.

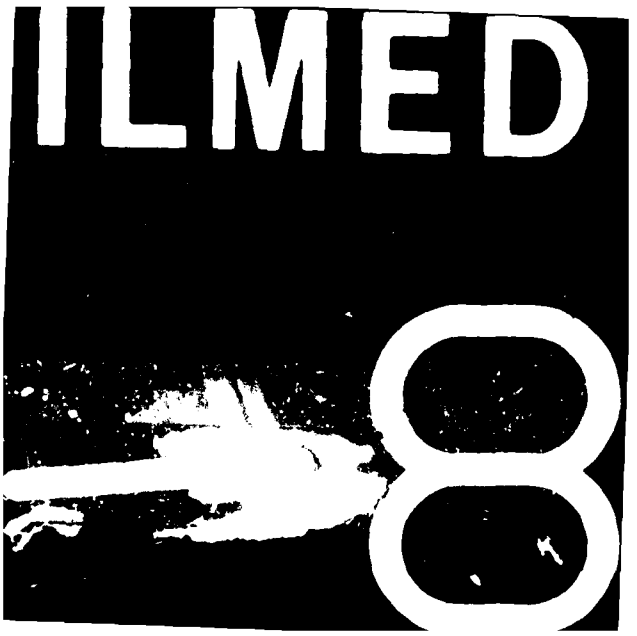
_____. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. Biometrika, 50, 129-134.

Wiener, N. (1949). The Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York, Wiley. Includes Levinson (1946).

Wiggins, R. A. and Robinson, E. A. (1965). Recursive solution to the multichannel filtering problem, Journal of Geophysical Research, 70, 1885-1891.

Wold, H. (1938). A Study in the Analysis of Stationary Time Series. Stockholm: Almquist and Wiksell.

Yule, G. U. (1927). "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," Philosophical Trans. Royal Soc. London, Series A, 226, 267-298.


ILMED
8